

AQS Data Mart Direct Interface User Instructions

Version 1.0, January 09, 2008

The AQS Data Mart “Direct Interface” is an application you can use to query the database. You can access it by following the link from the Data Mart Access page (or here is the link: <http://www.epa.gov/ttn/airs/aqsdatamart/access/interface.htm>.)

Access and Setup

System requirements – to be able to run the application you need a web browser with the Java Runtime Engine version 1.5 or later.

You will also need a CDX Node account for the Data Mart. To get this, send an email request for a “Node account to access the AQS Data Mart” (as the subject) to the EPA Node Helpdesk at epacdx@csc.com. Include the following information:

- Your name
- Address (physical)
- Email
- Phone
- Supervisor’s name
- Supervisor’s email
- Supervisor’s phone

You will be sent a user ID (your email address) and a password.

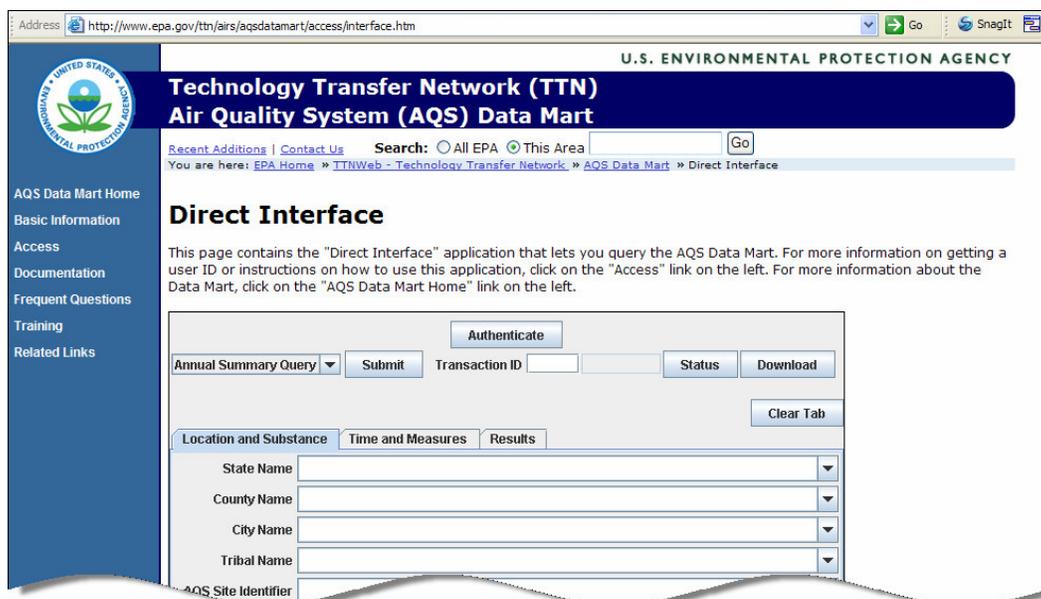
When you first access this page, you will probably get a security message that looks like this:



You may safely click “Run” to allow the application to open (Hermes is the name of an EPA server). If you want to avoid getting this message each time you open the application, check the “Always...” box before clicking “Run”.

Logging On

When you have successfully accessed the application, your browser window should look something like this:



The first step is to click on the “Authenticate” button. (Note, **you may need click on the application once to “activate” it within the browser**, so the first click on the “Authenticate” button may not have really pushed the button. If in doubt, click it again.)

A logon window will pop up:



Enter your User ID and Password and click “Connect”.

If you get an error message during the authentication step, either your account is invalid or you mistyped your User ID or Password. Also note that your logon is valid for only 30 minutes, so if you get a “token error” you must authenticate again.

Using the Direct Interface

Here is what the interface looks like when it starts up.

The screenshot shows the top portion of the Direct Interface. At the top center is an 'Authenticate' button. Below it, on the left, is a dropdown menu currently set to 'Annual Summary Query'. To its right is a 'Submit' button, followed by a 'Transaction ID' input field, a 'Status' button, and a 'Download' button. On the far right is a 'Clear Tab' button. Below these elements are three tabs: 'Location and Substance' (which is selected), 'Time and Measures', and 'Results'. Under the 'Location and Substance' tab, there are several input fields: 'State Name', 'County Name', 'City Name', 'Tribal Name', 'AQS Site Identifier', and 'EPA Region Number', all with dropdown arrows. Below these are 'Location' fields for 'Latitude' and 'Longitude', each with 'Min.' and 'Max.' sub-inputs. At the bottom of this section are 'Substance Type', 'Substance Name', and 'Monitor Type', all with dropdown arrows.

There are 4 steps to getting data by using the Direct Interface:

1. **Choose** the type of query you want to run
2. **Build** your query by entering selection criteria
3. **Submit** the query for execution
4. **Download** the results

Step 1 – Choose the type of query you want to run

At the top left of the screen is a pull down menu that lists the 5 types of query available.

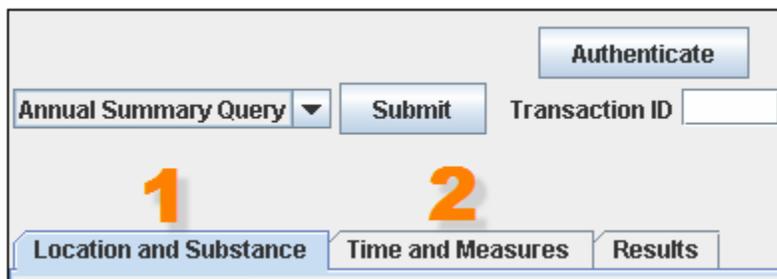
This close-up shows the 'Annual Summary Query' dropdown menu expanded. The menu lists five options: 'Annual Summary Query', 'Monitor Query', 'Raw Data Query', 'Sites by Threshold Query', and 'Values Query'. An orange arrow points down to the dropdown arrow, and another orange arrow points to the 'Values Query' option, which is highlighted with a blue background. A mouse cursor is visible over the 'Values Query' option. The background shows the 'Authenticate' button, 'Submit' button, 'Transaction ID' field, and the 'Time and Measures' and 'Results' tabs.

Each query type differs in the data that is returned and the format it is in. **You will almost always want to run the Values Query.** It allows you to select any of the measured data available in the Data Mart and returns it in the most easy to read format. See Appendix A for a list of the statistics (values) this query can return. Here is a description of all of the query types if you think you might want something different.

QUERY TYPE	DESCRIPTION
Values	Recommended. This is the most versatile and efficient query. It lets the user select from the most commonly requested values (raw, NAAQS averages, daily summary, and annual summary – see Appendix A) and returns them in a latitude/longitude oriented schema. Only one kind of value/statistic can be returned with each query.
Monitor	Returns a list of sites and monitors with their metadata (including latitude and longitude and descriptions of the monitoring site and equipment) in the AQS Submission schema.
Annual Summary	Returns all annual aggregate statistics for the selected monitors. It corresponds to the AQS “Annual Summary” report. It uses the AQS Summary schema.
Raw Data	This returns raw and NAAQS average data in the AQS Submission schema format. This format is indexed by tribe or state + county and does not contain latitude and longitude. It is provided primarily for the convenience of those familiar with the AQS system.
Sites by Threshold	Returns a list of sites and monitors with their metadata (including latitude and longitude) in the AQS Submission schema format that meet a user selected threshold. For example, monitors in Idaho with an annual maximum greater than 10.

Step 2 – Build your query by entering selection criteria

There are two tabs for entering your selection criteria: “Location and Substance” and “Time and Measures”. The screen defaults to the first.



Using the pulldowns, populate the first tab with valid combinations (see Appendix B for valid geographic combinations). Note, every time you request a pull down menu, there may be a delay as the database is actually being queried to get the requested list. (Sometimes the first time you use a pulldown menu in the application you will get an error. Close the error box and try again).

For example, let's select Jefferson County, Alabama and Criteria Pollutants. Here's what the screen will look like, with the used pulldowns highlighted.

The screenshot shows a web application interface with a search form. At the top, there is an 'Authenticate' button. Below it, there is a dropdown menu for 'Annual Summary Query', a 'Submit' button, a 'Transaction ID' input field, a 'Status' button, and a 'Download' button. A 'Clear Tab' button is located on the right side. The form has three tabs: 'Location and Substance', 'Time and Measures', and 'Results'. The 'Location and Substance' tab is active. It contains several dropdown menus: 'State Name' (highlighted with an orange double-headed arrow), 'County Name' (highlighted with an orange double-headed arrow), 'City Name', 'Tribal Name', 'AQS Site Identifier', and 'EPA Region Number'. Below these are 'Location' fields for 'Latitude' and 'Longitude', each with 'Min.' and 'Max.' sub-fields. The 'Substance Type' dropdown is also highlighted with an orange double-headed arrow. Below it are 'Substance Name' and 'Monitor Type' dropdown menus.

Note that every selection you make will limit the selections below it. That is, once you select a state, only counties within that state are listed and once you select a county, only AQS Site IDs within that county are listed. Also, for substances (pollutants), once you select a Substance Type, only the Substance Names within that type are listed. This is a handy way to shorten lists that would otherwise be very long.

You must select at least one of the geographic selections and at least one of the substance selections. Monitor Type is optional.

Once you have selected geography and substance click on the "Time and Measures" tab to complete the entry of you selection criteria. If we select the week beginning May 10, 2005 by entering the begin and end date fields (note the date format is YYYYMMDD) and select a value (Statistic Name) to be returned of Daily Maximum Sample Measurement, the screen will look like the one below:

The screenshot shows a web application interface for submitting a query. At the top, there is an 'Authenticate' button. Below it, a dropdown menu is set to 'Annual Summary Query', followed by a 'Submit' button, a 'Transaction ID' input field, a 'Status' button, and a 'Download' button. A 'Clear Tab' button is located on the right. The main form has three tabs: 'Location and Substan...', 'Time and Measures' (which is selected and circled in red), and 'Results'. The 'Time and Measures' tab contains several fields: 'Method Type' (dropdown), 'Method ID' (dropdown), 'Sample Duration' (input), 'Begin Date' (20050510, format yyyyymmdd), 'Begin Time' (input, format hh:mm:ss), 'End Date' (20050517, format yyyyymmdd), 'End Time' (input, format hh:mm:ss), 'Time Zone Type' (dropdown), 'Data Source' (dropdown), 'Statistic Name' (Daily Maximum Sample Measurements - Maximum sample measurements, by monit..., dropdown), 'Relational Operator' (dropdown), and 'Threshold Value' (input). At the bottom, there is an 'Include' section with checkboxes: 'All POCs' (checked), 'Event Data' (checked), 'Monitor Type' (unchecked), 'Method' (unchecked), 'Duration' (unchecked), 'Locational Metadata' (unchecked), and 'Measurement Metadata' (unchecked).

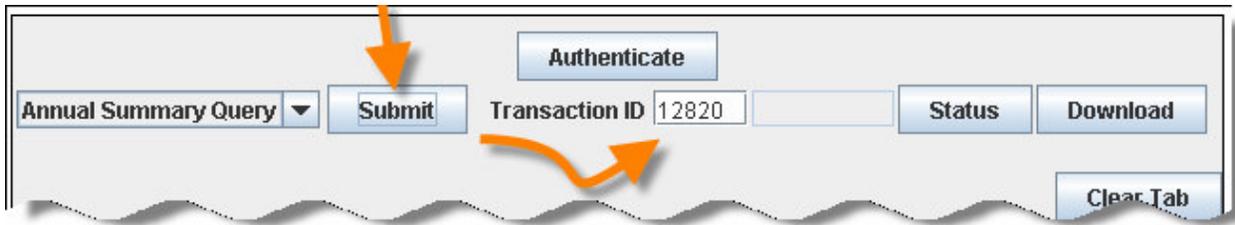
Note, we have selected only the mandatory fields. Method Type, Method ID, Sample Duration, and Times are optional and will default to “All” if not selected. Time Zone Type (of the returned data) will default to Local Standard Time if not selected (the other option is Greenwich Mean Time). Data Source defaults to AQS.

Some selection fields will be grayed-out if they are not valid selections for this query type.

The bottom of the screen lists optional data to include in the results and two are selected by default. “All POCs” indicates that, if there are multiple monitors for the same substance at a single site, you would like data from all monitors returned. “Event Data” indicates that data flagged by the submitter as not appropriate for regulatory use due to an exceptional event should be included in returned set. The other check boxes (grayed-out for the Daily Maximum selection) indicate additional metadata that can be returned with the data. So, the first two (selected by default) apply to measured data and the remaining five apply to metadata.

Step 3 – Submit the Query for Execution

Now that you’ve completed entering the selection criteria, send the query off for processing by pressing the “Submit” button back at the top of the screen. When you do this (and the database receives it) a Transaction ID will be assigned to your query:



Step 4 – Download the Results of your Query

There are two ways to download the results of your query.

The easiest way is to wait for an email from “aqsdatabmart@EPA” that tells you your data is ready for you to download, and click on the link in the body of the email.

The subject will be: AQS Data Mart report AQDM_RR_XXXXX.xml is ready

Where “XXXXX” will be the Transaction ID assigned when you clicked the “Submit” button. The body of the email will be something like:

Your AQS Data Mart report has completed processing and may be retrieved at the link below or by using the client to download the results.

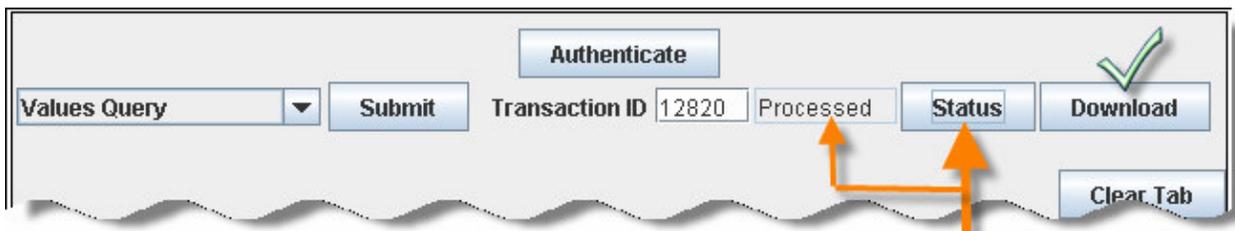
https://iasint.rtpnc.epa.gov/AQDM/AQDM_RR_XXXXX.xml

Note, the number at the end of the filename (12820) is the transaction ID you were issued when the job was submitted. This file has a size of 186,556 bytes.

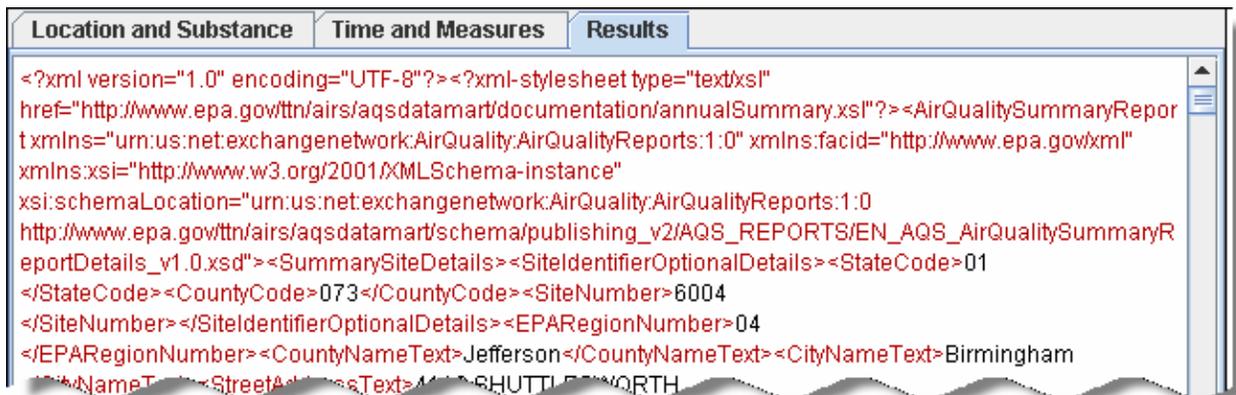
You can simply click on the link in the email to open the XML file in your browser. Note that XML files are displayed in browsers using stylesheets for the layout and format. Our stylesheets do not display properly in Mozilla at this time.

If the file is very large (or if you want to save it without opening it in a browser) and you are using a web enabled email, you can right click on the link and select “Save target as...” and you will be able to save the file directly to your local disk.

The other way to download your results is to use the interface to request the status of the query from the database, and when the status is “Processed”, click the “Download” button. (Keep clicking the “Status” button until the job is “Processed”).



After you click download, a dialog box asking you where to save the file will open. **Be sure to save the file with a .xml extension.** After the file is saved it will also be displayed in the “Results” tab of the Interface:



Note that, due to EPA server memory limitations, **files over 50MB cannot be downloaded through the Interface** and must be downloaded via the link in the email.

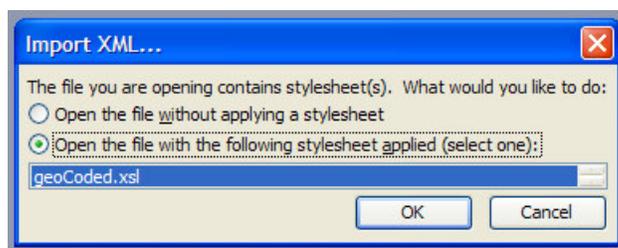
If the Status comes back as “Failed” you can click on “Download” to see the error message / reason for the failure.

About the Output and XML

All of the data output by the Data Mart is in XML format. This is a format where each value is surrounded by tags that indicate what the value represents. XML is very handy when a machine is reading the data, but can be off-putting when a person is trying to read the data. XML is becoming easier to deal with, and please keep in mind:

- It is easily formatted into a human readable format using something called a “stylesheet”
- You probably have an XML stylesheet processor on your computer
- Almost all web browsers are stylesheet processors (IE, Mozilla, etc.)
- We have provided a stylesheet for you to use
- The stylesheet is on our web page and a link to it is already in the XML file
- For the extremely XML adverse, we plan to provide a tool to convert Data Mart XML to CSV (comma separated values) text files later in 2008.

For example, you can open an AQS Data Mart XML file in Microsoft Excel. When Excel realizes it has an XML file it will provide a dialog box, choose the option that says: “Open with the following stylsheet applied”.



If, for some reason, the stylesheet is not available, you will get a different dialog box. In this case select to open the file “As read-only workbook”. It will not look as good as with the stylesheet, but will be readable.

Here is a list of the fields in the output of a Values Query and their meanings.

FIELD	DESCRIPTION
Latitude Measure	The monitoring site's angular distance north or south of the equator measured in decimal degrees.
Longitude Measure	The monitoring site's angular distance east of the prime meridian measured in decimal degrees.
Parameter	The name assigned in AQS to the parameter measured by the monitor. Parameters may be pollutants or non-pollutants (this term has the same meaning as “substance”).
Data Source Reference ID	The ID number of the site in the AQS system. This is for cross reference if you have data from another query or system.
Date	The calendar date. All annual summary data is assigned to January 01.
Time	The time of day on a 24-hour clock. All daily summary data is assigned to midnight (00:00).
Time Zone	The time zone where the monitor resides.
Year & Day in Year	An alternate way of expressing the date that is more useful for indexing and sorting. Not yet populated.
Statistic Name	The name of a summary value or statistic requested in the query.
Measure Value	The measured or calculated value in the database for the requested statistic.
Measure Unit Name	The units of measure of the measure value field.
Measurement Uncertainty	The total uncertainty associated with a reported measurement as indicated by the reporting agency.
Duration Description	The length of time that air passes through the monitoring device before it is analyzed (measured). So, it represents an averaging period in the atmosphere (for example, a 24-hour sample duration draws ambient air over a collection filter for 24 straight hours). For continuous monitors, it can represent an averaging time of many samples (for example, a 1-hour value may be the average of four one-minute samples collected during each quarter of the hour).
Observation Count	The number of observations (sample values) reported during the time in question (day or year).
Method Description	The text that describes the process and/or tools that manage storage, disposal, treatment, and other handling protocols designed for and/or used in taking the sample.

Monitor Type	An administrative classification for the monitor roughly indicating the monitoring network to which the monitor belongs. A monitor may have more than one type, and there is usually only one space to indicate this information in reports, so a hierarchy is used to select the most prevalent type for inclusion.
MDL	Method detectable limit (MDL) is the minimum detectable level of a sample concentration defined for the monitor and method.
Qualifier Description	Sample values may have qualifiers that indicate why they are missing or that they are out of the ordinary. Types of qualifiers are: null data, exceptional event, natural events, and quality assurance.
Horizontal Datum	The edition of North American Datum used as the basis for determining the site coordinates as listed by the agency reporting data to AQS.
Horizontal Accuracy Measure Value	Description of the accuracy of the site coordinates as listed by the agency reporting data to AQS.
Horizontal Accuracy Measure Unit	The units of measure of the horizontal accuracy.
Vertical Measure Value	The elevation above or below mean sea level (MSL) of the site as listed by the agency reporting data to AQS.
Vertical Measure Unit	The units of measure of the vertical measure.
Vertical Accuracy Measure Value	Description of the accuracy of the elevation as listed by the agency reporting data to AQS.
Vertical Accuracy Measure Unit	The units of measure of the vertical accuracy.

How to Get Help

If you have trouble accessing or using the AQS Data Mart Direct interface or interpreting the data you've received, you may get help by contacting using one of the following methods.

1. Call the EPA help desk at 1-866-411-4372 and tell them you need help with the AQS Data Mart
2. Email the EPA help desk at epacallcenter@epa.gov
3. Email the Data Mart system managers at: aqsdatamart@epa.gov

Appendix A – Statistics Available from the Values Query

Below is a list of the items on the “Statistics” pull-down menu for the Values Query and their meanings. The list is broken into three tables for display in this document.

The first table represents “raw” data. Raw data is both actual sample measurements and calculated National Ambient Air Quality Standard (NAAQS) average values.

STATISTIC	DESCRIPTION
Sample Measurements in Standard Units	Air quality measurements converted from the reported units of measure to the standard units of measure.
Sample Measurements in Reported Units	Air quality measurements in the units of measure reported by the submitter.
8-Hour Running Averages	Averages obtained from the sum of the hourly measurements in NAAQS 8-Hour intervals, divided by the number of measurements taken during the interval.
24-Hour Values	Measurement values for 24-hour duration samples <u>and</u> averages obtained from the sum of the hourly measurements in NAAQS 24-hour intervals divided by the corresponding number of measurements in those intervals.

The second group of statistics is daily aggregate values, and returns one value per day per monitor.

STATISTIC	DESCRIPTION
Daily Maximum Sample Measurements	Maximum sample measurement for each day, by monitor.
Daily Maximum 8-Hour Running Averages	Maximum 8-hour average value for each day, by monitor.
Daily Means of Sample Measurements	Measures of central tendency obtained from the sum of sample measurement values divided by the number of values that comprise that sum, by monitor and day.
Daily Percents of Sample Measurements	Ratios of reported sample measurements to possible (scheduled) sample measurements, by monitor and day.
Daily Percents of 8-Hour Running Averages	Ratios of 8-hour running averages to scheduled NAAQS 8-hour intervals, by monitor and day.
Daily Percents of 24-Hour Values	Ratios of 24-hour values to scheduled 24-hour values, by monitor and day.
Daily AQIs	Air Quality Index values, by monitor and day.

The final group of statistics represents annual aggregate values, and returns one value per year per monitor.

STATISTIC	DESCRIPTION
Annual Percents of Sample Measurements	Ratios of reported sample measurements to possible (scheduled) sample measurements, by monitor and year.
Annual Percents of 8-Hour Running Averages	Ratios of 8-hour running averages to scheduled NAAQS 8-hour intervals, by monitor and year.
Annual Percents of 24-Hour Values	Ratios of 24-hour values to scheduled 24-hour values, by monitor and year.
Annual Means of Sample Measurements	Measures of central tendency obtained from the sum of sample measurement values divided by the number of values that comprise that sum, by monitor and year.
Annual Means of 8-Hour Running Averages	Measures of central tendency obtained from the sum of 8-hour averages divided by the number of averages that comprise that sum, by monitor and year.
Annual Maximum Sample Measurements	Maximum sample measurements, by monitor and year.
Annual Maximum 8-Hour Running Averages	Maximum 8-hour running averages, by monitor and year.
Annual Maximum 24-Hour Values	Maximum 24-hour values, by monitor and year.
Annual 4th Maximum 8-Hour Running Averages	4th maximum 8-hour running averages, by monitor and year.
Annual 98th Percentiles of 24-Hour Values	24-hour values where 98% of the other 24-hour values in the same monitor and year value set are less than or equal to them.
Annual Counts of Sample Measurements above the Standard	Counts of sample measurements that exceed the applicable primary, (i.e., health-related) standard, by monitor and year.
Annual Counts of 8-Hour Running Averages above the Standard	Counts of 8-hour running averages that exceed the applicable primary, (i.e., health-related) standard, by monitor and year.
Annual Counts of 24-Hour Values above the Standard	Counts of 24-hour values that exceed the applicable primary, (i.e., health-related) standard, by monitor and year.
Annual Counts of Sample Measurements with Exceptions	Counts of sample measurements that were flagged as being affected by exceptional events, by monitor and year.

Appendix B – Allowable Geographic Selection Combinations

The following single and combinations of geographic selection are allowed. Note, there is no national selection allowed at this time. If you need data for the entire nation, we request that you run 10 queries, one for each EPA region. (We do this to avoid overloading the servers.)

State
State + County
State + City
State + County + AQS Site ID
Tribe
Tribe + AQS Site ID
EPA Region Number
Latitude / Longitude Box

The screenshot shows a web application interface with three tabs: "Location and Substance", "Time and Measures", and "Results". The "Location and Substance" tab is active. It contains several dropdown menus and input fields:

- State Name:** A dropdown menu with an orange dot on the left and four vertical orange lines extending downwards.
- County Name:** A dropdown menu with two vertical orange lines extending downwards.
- City Name:** A dropdown menu with one vertical orange line extending downwards.
- Tribal Name:** A dropdown menu with two vertical orange lines extending downwards.
- AQS Site Identifier:** A dropdown menu with one vertical orange line extending downwards.
- EPA Region Number:** A dropdown menu with an orange dot on the left.

Below the dropdown menus, there are input fields for "Location":

- Latitude:** Two input boxes labeled "Min." and "Max." with an orange dot on the right.
- Longitude:** Two input boxes labeled "Min." and "Max." with an orange dot on the right.

Appendix C – Miscellaneous Technical Notes

This appendix contains technical notes and solutions to problems that some users have encountered.



If you get an "access denied" error when you click on a link the Data Mart has emailed you and you are running Internet Explorer (IE) version 6 or 7, the file will properly display if you make the following change to you options:

1. From IE, select the menu pick "Tools / Internet Options"
2. Click on the Security tab.
3. Click the "Custom Level" button.
4. Scroll down to the "Miscellaneous" section.
5. Under "Access data sources across domains", select the "Prompt" checkbox.
6. Click OK until you exit the security dialogue.
7. Restart IE.



Your output is kept for approximately 14 days on the EPA servers. If you need to download the same file again within that window, you can click on the email link or enter your Transaction ID on the Interface screen and select "Download".